SSRG International Journal of Recent Engineering Science
ISSN: 2349–7157 / https://doi.org/10.14445/23497157/IJRES-V10I3P104

Volume 10 Issue 3, 23-32, May-Jun 2023

Original Article

Financial Fraud Detection using Machine Learning : Credit Card Fraud

Odeajo Israel[1], Akinmoluwa Oluseye[2], Sharon Ojo[3], Otesanya Temitope Deborah[4]

[1]*Machine Learning Engineer, West Africa Union University, Republic of Benin.*
[2]*Lead City University, Ibadan, Nigeria.*
[3,4]*University of Ibadan, Nigeria.*

[1]*Corresponding Author : isrealodeajo@gmail.com*

Received: 03 April 2023 Revised: 05 May 2023 Accepted: 16 May 2023 Published: 31 May 2023

Abstract *- Instances of credit card fraud occur with great frequency and often lead to serious financial losses. The volume of online transactions has experienced significant growth, with a substantial proportion of those transactions being attributed to credit card transactions made online. Hence, credit card fraud detection applications are highly valued and in demand by banking institutions and financial institutions. Fraudulent transactions can manifest in diverse forms and can be classified into distinct categories. This research centers on cases of fraudulent activity from open-source data from kaggle.com. Fraudulent activities are examined by employing a sequence of machine learning models, and the optimal approach is determined through an extensive analysis process. We used three algorithms, namely the random forest algorithm, the Decision Tree classifier algorithm, linear regression and three sampling techniques in order to balance the dataset. We also used twelve (12) different models for the prediction of credit card fraud. The evaluation offers a comprehensive guide for the selection of an ideal algorithm based on the nature of fraudulent activities. Additionally, we demonstrate the evaluation process using a suitable metric for performance measurement. The twelve models were compared, and the best model, with an accuracy of 97.4%, was a Random Forest Classifier developed using the SMOTE sampling technique after hyperparameter tuning.*

Keywords *- Algorithm, Credit card fraud, Decision tree, Fraudulent transactions, Random forest.*

1. Introduction

The rapid increase in the usage of internet banking systems for cash transactions, bill payments, and shopping has made life easier for most people. However, despite the great benefit of online transactions, financial fraud is causing a huge problem for users. Fraud happens when a third party operates as though they are the real customer by bypassing the bank's security measures and passing themselves off as that customer. Financial fraud is an epidemic that only worsens and has far-reaching effects on the financial sector.

According to [1], In 2019, Belgium witnessed the reporting of over 12,000 instances of internet banking fraud. This work represents the peak value attained since 2006. Since the inception of digital payments, the payments industry has endeavored to establish a secure environment for financial transactions. Phishing emails are the predominant form of fraudulent activity in the internet or online banking realm. This method involves using deceptive tactics to lure individuals into divulging confidential financial information [2].

The credit card, which is widely utilized in financial transactions, is intended to facilitate purchases of various commodities, including but not limited to fuel, groceries, electronic devices, travel expenses, and retail bills, particularly in situations where cash on hand are not readily accessible [3]. According to recent data, digital wallets, credit, and debit cards have emerged as the most frequently utilized means of payment for e-commerce transactions on a global scale in 2021 [4]. It was approximated that digital and mobile wallet payments constituted approximately 50% of global online transactions. According to recent market research, online wallets dominated the Asia-Pacific Region's e-commerce payment landscape, constituting nearly 70% of the market share [4].

The global losses incurred due to card fraud in 2021 amounted to $32.34 billion, representing a 14% surge from the previous year's losses of $28.43 billion. The aggregate amount of fraud losses in the United States for the year 2021 amounted to $11.91 billion, signifying an 18% surge from the $10.09 billion recorded in 2020 [5]. The incidence of card not present fraud has risen above that of point of sale fraud by 81%. American banks, businesses, and cardholders report credit card fraud daily. According to a recent report [6], payment card fraud losses worldwide are primarily attributed to accounting for 38.6% of the total. Various categories can be

used to classify credit card fraud. Two primary categories of fraud that can be identified within a given set of transactions are Card-Not-Present (CNP) fraud and Card-Present (CP) fraud [7-10].

According to recent data, Nigerian financial services firms incurred a loss of ₦5.2 billion due to fraudulent activities within the period spanning January to September of the year 2020 [11]. The majority of this financial deficit was incurred during the period spanning from July to September 2020, during which companies experienced a loss of as much as ₦3.36 billion. There was a significant surge of 510% in the amount lost to fraudsters during the same period in 2019, compared to the current period, which amounted to ₦550 million [12]. According to the 2021 fraud report released by the NIBSS, the total fraud attempts in Nigeria increased by 187% between 2019 and 2020. The top sources of fraud in 2020 were found to be the web, accounting for 47% of transactions, followed by mobile at 36%, with ATM terminals and POS terminals accounting for 9% and 7% of transactions, respectively.[13].

The machine learning domain can be broadly categorised into two primary classifications, specifically supervised learning and unsupervised learning [14-16]. Machine learning algorithms involve using historical data to train models that can predict fraudulent transactions. The utilisation of data analytics has gained significant traction in the realm of financial fraud detection owing to its capacity to analyse vast quantities of data and discern complex patterns that are difficult to detect through conventional methods. Several machine learning algorithms are commonly employed in the context of fraud detection, such as logistic regression, decision trees, random forests, support vector machines, and neural networks [17, 18].

The objective of our research is to investigate the characteristics of Card Not Present (CNP) fraud and to present a machine learning-based approach for its detection. The present study employs two primary methods of data analysis, namely categorical and numerical analysis, to examine the data under investigation. This study involves the identification of the most suitable algorithms for detecting fraudulent patterns in credit card transactions by conducting a comprehensive analysis of various machine-learning techniques. The effectiveness of the algorithms will be evaluated using a measure of performance for detecting fraudulent credit card transactions.

2. Literature Review

2.1. Financial Fraud

Various classifications of financial fraud exist, with credit or debit card fraud being the most prevalent. Figure 1 illustrates shows four distinct categories of Credit Card Fraud, namely card-not-present transactions, skimming, phishing, and lost or stolen cards [32].

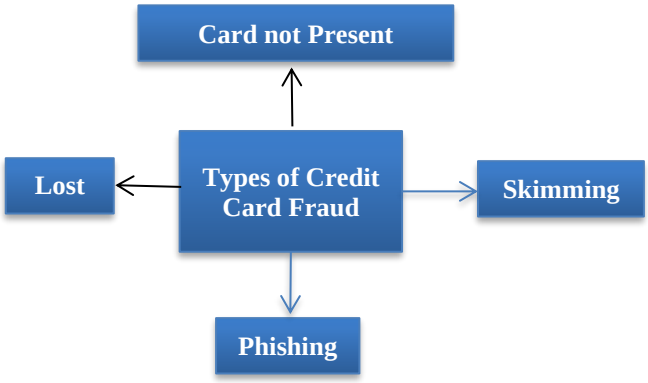

Fig. 1 Types of credit card fraud

Card Not Present (CNP) fraud involves fraudulent attempts to deceive the system by impersonating another individual [32]. The employment of mail and web channels has emerged as significant avenues for perpetrating fraudulent activities against merchants engaged in the sale and shipment of merchandise. This phenomenon has detrimentally impacted authentic mail orders and web-based merchants. Skimming involves the acquisition of personal information pertaining to an individual's credit card that has been utilised in an otherwise routine transaction[21]. A skimmer is utilised to surreptitiously acquire and retain substantial quantities of personal data belonging to unsuspecting individuals. Phishing is a fraudulent activity that involves the use of various tactics by scammers to deceive users into divulging their card information. These tactics may include creating counterfeit websites that mimic those of legitimate banks or payment systems. In instances where a card is stolen or lost, there exists a possibility for an unauthorised individual to conduct transactions before the cardholder takes the necessary measures to block the card [32]. However, works have been done by different authors to checkmate this financial fraud class.

Used GA algorithm for the purpose of feature selection in a machine learning-based credit card fraud detection system[21]. The authors combined the RF, DT, ANN, NB, and LR with a GA-based feature selection method for this study. The RF was incorporated into the GA's fitness function for optimal performance. Five optimal feature vectors were generated after the genetic algorithm was utilised to analyse a dataset consisting of credit card transactions done by cardholders from Europe. They found that GA-RF (using v5) provided the highest accuracy overall. The GA-DT and other classification models exhibited exceptional performance with v1, achieving a 99.92% accuracy rate. The results of this investigation have exceeded the current techniques.

Implemented a deep learning algorithm to conduct a forensic detection of credit card fraud transactions using an LSTM model that combined several machine learning techniques[22]. From their results, LSTM-attention

algorithms can be used to conduct forensic credit card fraud detection with high accuracy and precision.

The study conducted [23] aimed to evaluate the effectiveness of naïve Bayes, k-nearest neighbour, and logistic regression algorithms on credit card fraud data with high skewness. The results indicated that the k-nearest neighbour outperformed both naïve Bayes and logistic regression in terms of performance. [24] employed various machine learning algorithms, namely Local Outlier Factor, Support Vector Machine, Logistic Regression, Decision Tree, and Random Forest, for the purpose of detecting credit card fraud. The algorithms of logistic regression, decision trees, and random forests demonstrated superior performance.

Proposes and conducts an investigation on seven hybrid machine learning models for the purpose of detecting fraudulent activities [25]. The study employs a real-world dataset. The formulated hybrid models comprised two distinct phases, wherein cutting-edge machine learning algorithms were initially employed to identify instances of credit card fraud. The study's results revealed that the hybrid Adaboost + LGBM model demonstrated superior performance compared to other models, thus earning the title of the champion model.

The study focused on investigating the utilisation of a machine learning methodology to identify instances of credit card fraud autonomously [26]. The researchers employed the Long Short-Term Memory-Recurrent Neural Network (LSTM-RNN) technique in the context of credit card fraud detection.

Implemented a machine learning algorithm and utilised three convolutional neural network architectures to enhance the efficacy of fraud detection [27]. The findings of their study demonstrated enhanced outcomes in terms of accuracy, f1-score, precision, and AUC Curves, with optimised values of 99.9%, 85.71%, 93%, and 98%, respectively. The model under consideration exhibits superior performance in comparison to the existing state-of-the-art machine learning and deep learning algorithms when applied to credit card detection problems.

The best performance was the C4.5 decision tree with an accuracy of 94.13% precision, with the least performance being naïve bayes with 65.6% precision. In a study conducted by [28], the efficacy of the Deep Convolutional Neural Network (DCNN) model was compared with that of Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF) for the purpose of credit card fraud detection. The results indicated that the DCNN model outperformed SVM in both speed and accuracy.

2.2. Theoretical Background

Below, we explain the classification algorithms that were used in this paper.

2.2.1. Logistic Regression

Logistic regression is a statistical technique utilised for binary classification. It involves the utilisation of a linear model. Therefore, it is utilised to conduct regression analysis on a set of variables. It is a technique for predicting data patterns with unambiguous or numerical parameters [32]. The process of determining probability involves the utilisation of a set of input vectors and a corresponding response variable that is dependent, with the application of logarithmic functions. Probability belongs to a specific class. For binary classification, the response variable is given as [32]:

$$\mathrm{Yi} = \begin{cases} 0 \\ 1 \end{cases} \tag{1}$$

Therefore, the formula for determining the classification of sample Xi into class one is expressed as follows::

$$\mathrm{P} = (\mathrm{y_i} = 1|\mathrm{X_I}) = \frac{\exp\,(\mathrm{W_0} + \mathrm{W^T X_i})}{1+\exp(\mathrm{W_0} + \mathrm{W^T X_i})} \tag{2}$$

Where W0 and W represent the standardisation parameters used in regression, W0 is the intercept, and W is the vector of coefficients [15].

2.2.2. Decision Tree

Decision trees employ a binary tree structure with successive nodes to facilitate data classification. As the sample progresses through the tree, it is determined by the creation of the respective node [29]. The tree is partitioned into subsets that are successively segregated into mutually exclusive subgroups until they are ultimately stored. [32] Decision tree is also known as a classification and regression tree.

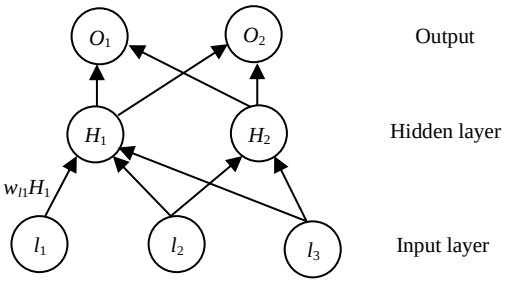

Fig. 2 Binary decision tree

Source: Popat & Chaudhary, 2018

If a target is a classification outcome taking on values 0,1, K-1, for node m,
Let

$$\mathrm{p_{mk}} = \frac{1}{\mathrm{n_m}} \sum_{\mathrm{y} \in \mathrm{Q_m}} \mathrm{I}\,(\mathrm{y} = \mathrm{k}) \tag{3}$$

be the proportion of class k observations in node m.

If m is a terminal node, common impurity measures are the following.

$$\mathrm{H(Q_m)} = \sum_{\mathrm{y} \in \mathrm{Q_m}} \mathrm{p_{mk}}(1 - \mathrm{p_{mk}}) \tag{4}$$

2.2.3. Random Forest Algorithm

The Random Forest algorithm is a commonly utilised supervised learning technique. This methodology is suitable for achieving both regression and classification objectives. It is important to note that this algorithm is predominantly employed for classification-related tasks. The Random Forest algorithm generates decision trees based on the sample data and obtains predictions from each sample data. The Random Forest algorithm is classified as an ensemble method. The algorithm exhibits superior performance compared to single decision trees due to its ability to mitigate over-fitting through result averaging.

The performance index, which solely approximates the Confidence Interval (CI) of the RF model, is given as

$$mg(x,y) = av_k I(h_k(x,\Theta_k) = y) - \max_{j \neq y} av_k I(h_k(x,\Theta_k) = j) \quad (5)$$

where $I(.)$ denotes an indicator function, and $av(.)$, the average value.

3. Materials and Methods

3.1. Data Description

The open-source data utilised in our study was sourced from kaggle.com (fraudTrain.csv dataset) [30]. The dataset comprises simulated credit card transactions, encompassing both genuine and fraudulent transactions, spanning from the 1st of January 2019 to the 31st of December 2020. The dataset encompasses credit card usage data from a sample of 1000 customers engaging in transactions with a collective of 800 merchants.

We imported required packages and plotted the distribution of each variable *'trans_date_trans_time'*, *'cc_num'*, *'merchant'*, *'category'*, *'amt'*, *'first'*, *'last'*, *'gender'*, *'street'*, *'city'*, *'state'*, *'zip'*, *'lat'*, *'long'*, *'city_pop'*, *'job'*, *'dob'*, *'trans_num'*, *'unix_time'*, *'merch_lat'*, *'merch_long'*, *'is_fraud*. Also, analyzed the transaction patterns of the customers as shown in Figure 3.

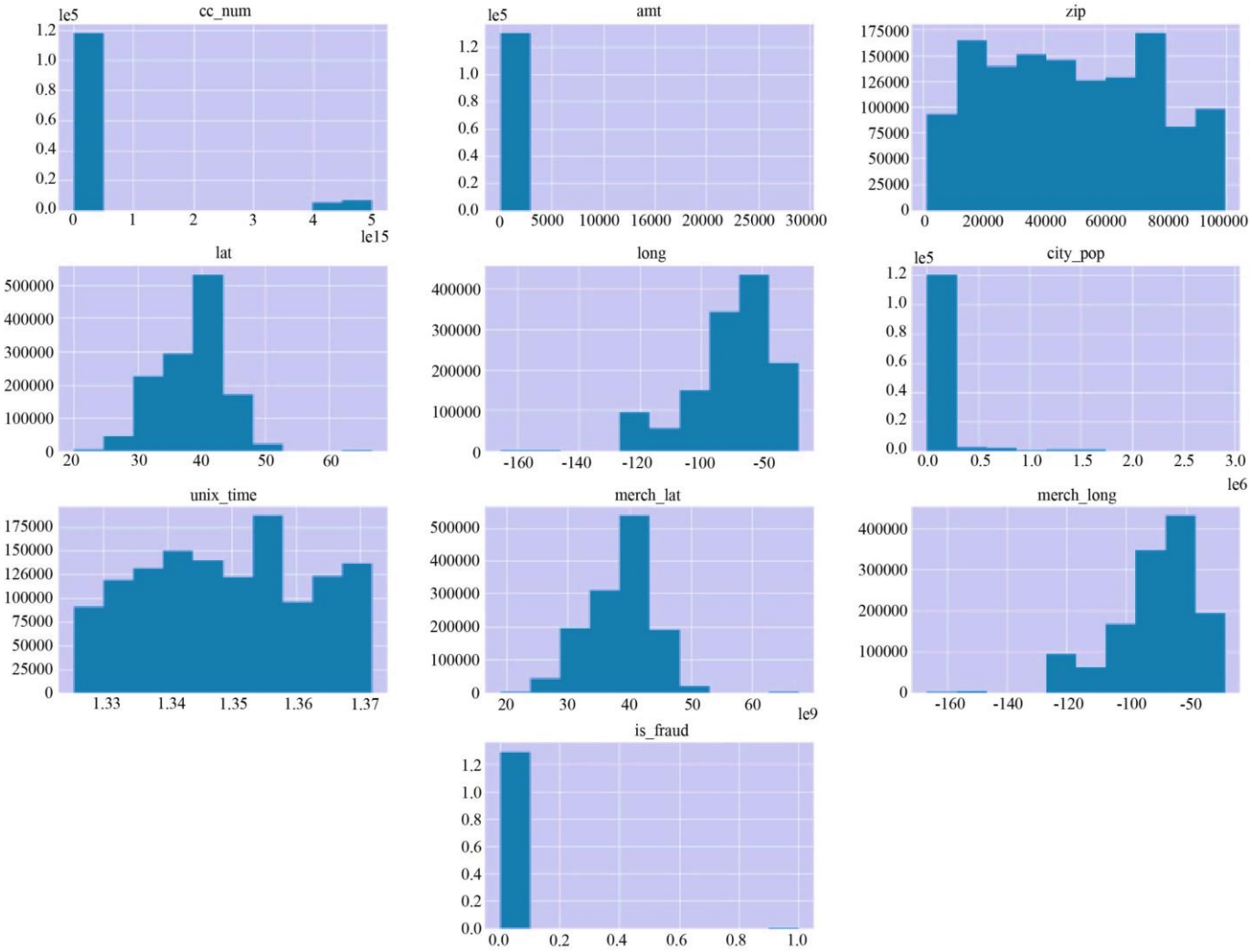

Fig. 3 Plot of the distribution of each variable

3.2. Feature Engineering

We extracted useful information from the "transaction date and transaction time" variable, converting transaction date and transaction time into date time, plotted the 'transanction hour' feature and plotted the 'transanction day of week' feature. We also grouped the year of the month and the number of transactions and plotted the year of the month vs the number of transactions.

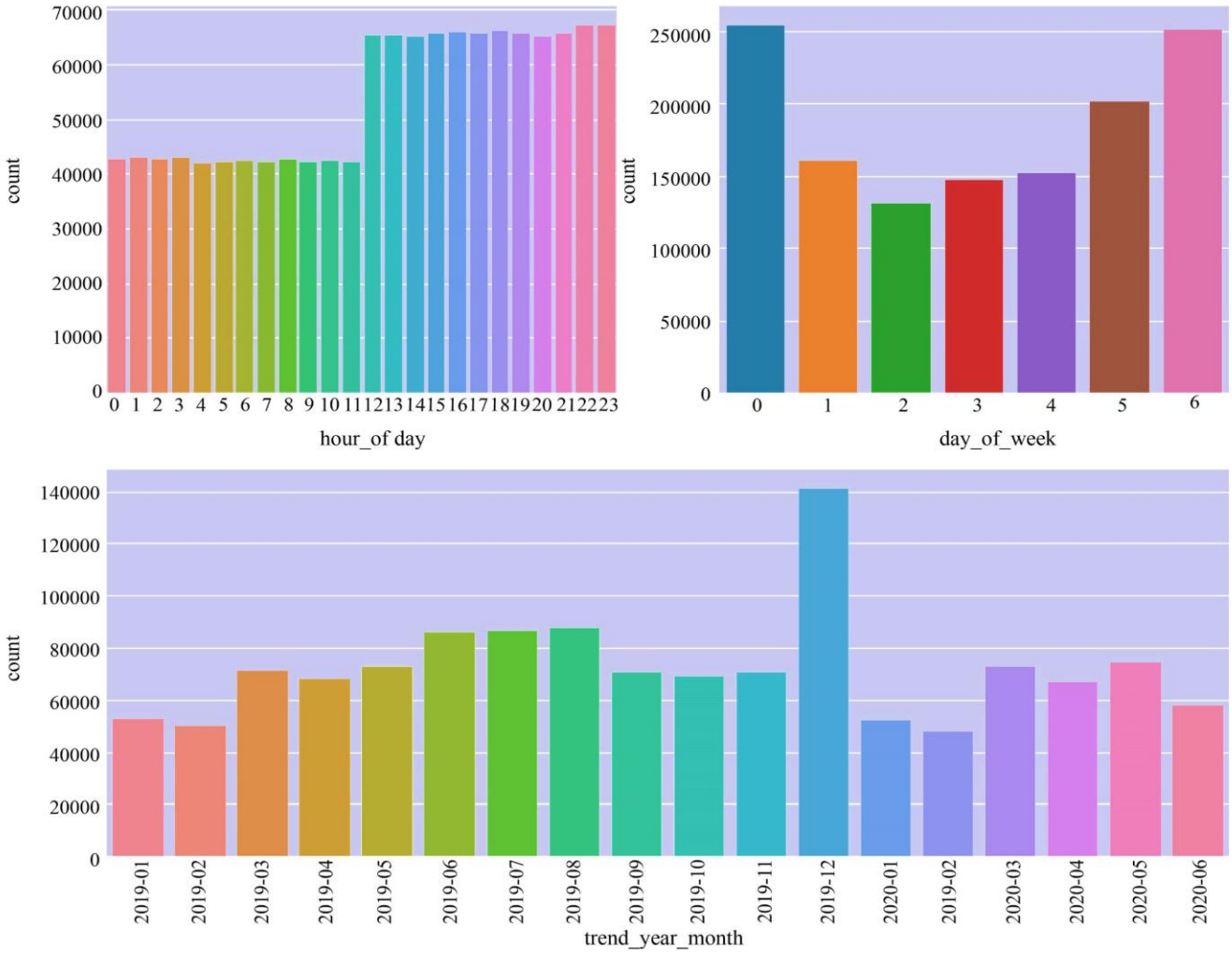

Fig. 4 Transaction hour, day of week, month of the year feature

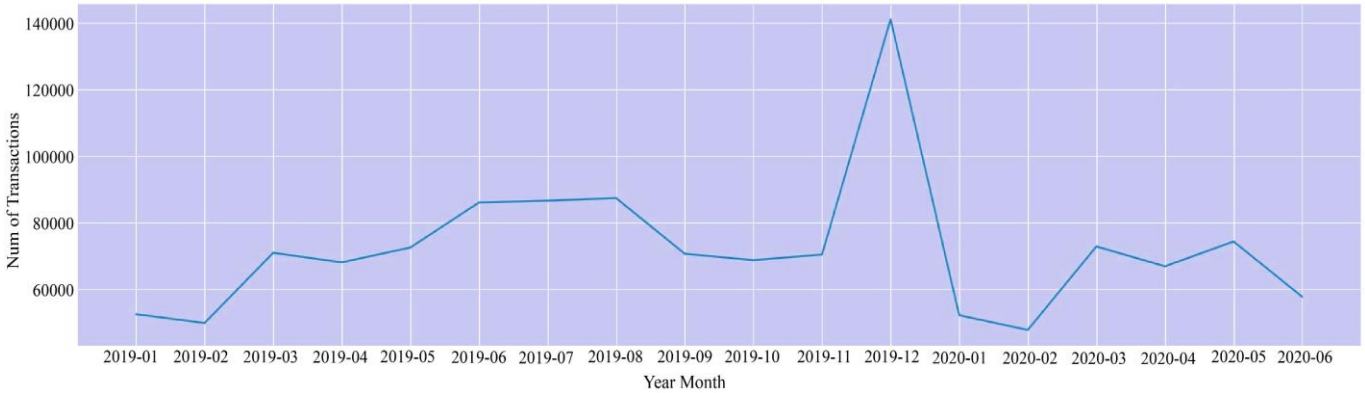

Fig. 5 Year of the month and number of transactions

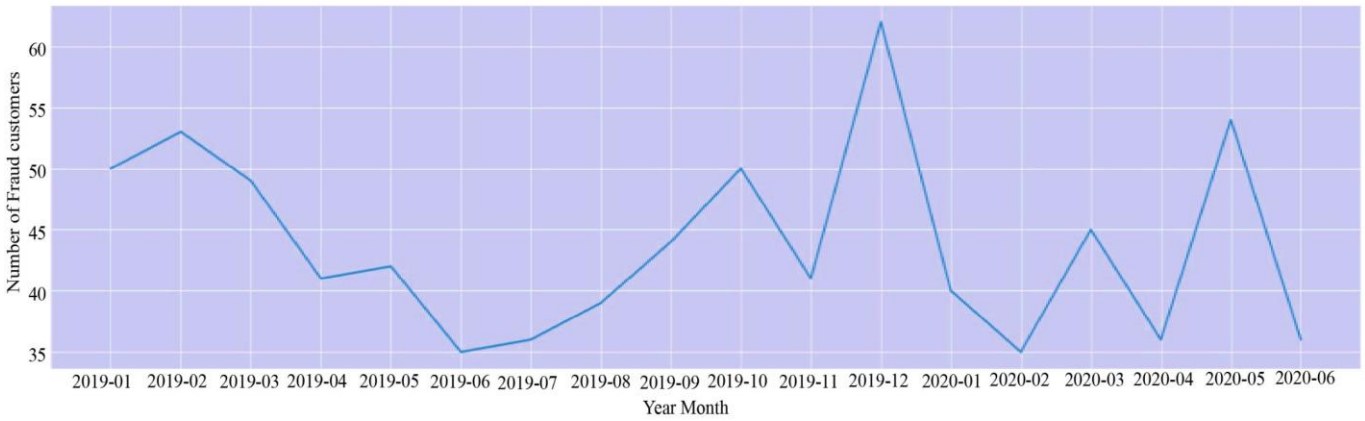

Fig. 6 Year and month, number of fraud transactions chart

Further, we grouped and plotted year and month, number of_fraud transactions and fraud customers. We grouped and plotted gender fraud distribution and created the age-fraud distribution as shown in Figures 7 and 8.

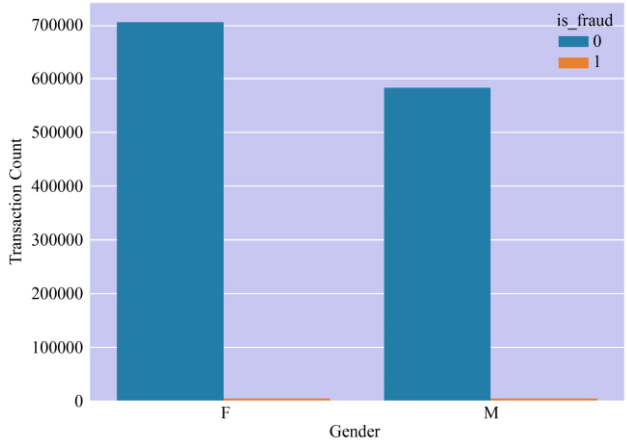

Fig. 7 Gender-Fraud distribution chart

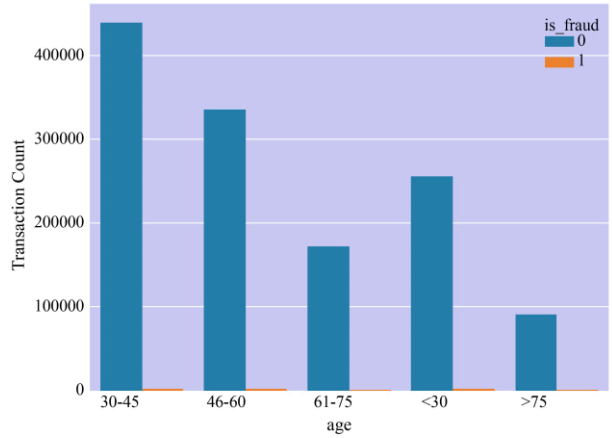

Fig. 8 Age-Fraud distribution chart

We identified jobs with more than one percent fraudulent transactions in the category feature, selected the top 5 most categories with the most fraud rates, and encoded the categorical variables from sklearn. Preprocessing and checking the correlations between the columns, plotting the correlation heatplot and getting the features with a correlation above 85%, as shown in Figure 9.

3.3. Model Building

The dataset was divided into two distinct subsets: a Training set of 70% and a Test set of 30% of the data. Our model's predictions used a variety of performance metrics, such as accuracy, F1, precision, and Recall matrix. We used two (2) different algorithms on the processed dataset and three (3) sampling techniques to balance the dataset. We also created and used eight (8) different models for the prediction of credit card fraud. The features 'zip', 'lat', 'long', 'city_pop', 'unix_time', 'merch_lat', and 'merch_long' have been assumed to provide no significant information in the model-building phase. Hence, they, along with the original features that have been encoded, have been dropped from the dataset.

From our findings, as shown in Table 1, the Logistic Regression - imbalance class' gave an accuracy of 99%, Logistic Regression with Random Under Sampling gave an accuracy of 84%, Logistic Regression - Random Over Sampling gave an accuracy of 84%, while Logistic Regression - SMOTE gave an accuracy of 84%

Also, the Decision Tree - imbalance class gave an accuracy of 99.8%, Decision Tree - Random Under Sampling gave an accuracy of 95%, Decision Tree - Random Over Sampling gave an accuracy of 94.6%, Decision Tree - SMOTE gave an accuracy of 94%

Random Forest - imbalance class gave an accuracy of 99.8%, Random Forest - Random Under Sampling gave an accuracy of 96.5%, Random Forest- Random Over Sampling gave an accuracy of 95.5%, Random Forest- SMOTE gave an accuracy of 94.8% and Random Forest - SMOTE [Hyperparameter Tuned] 97.4%.

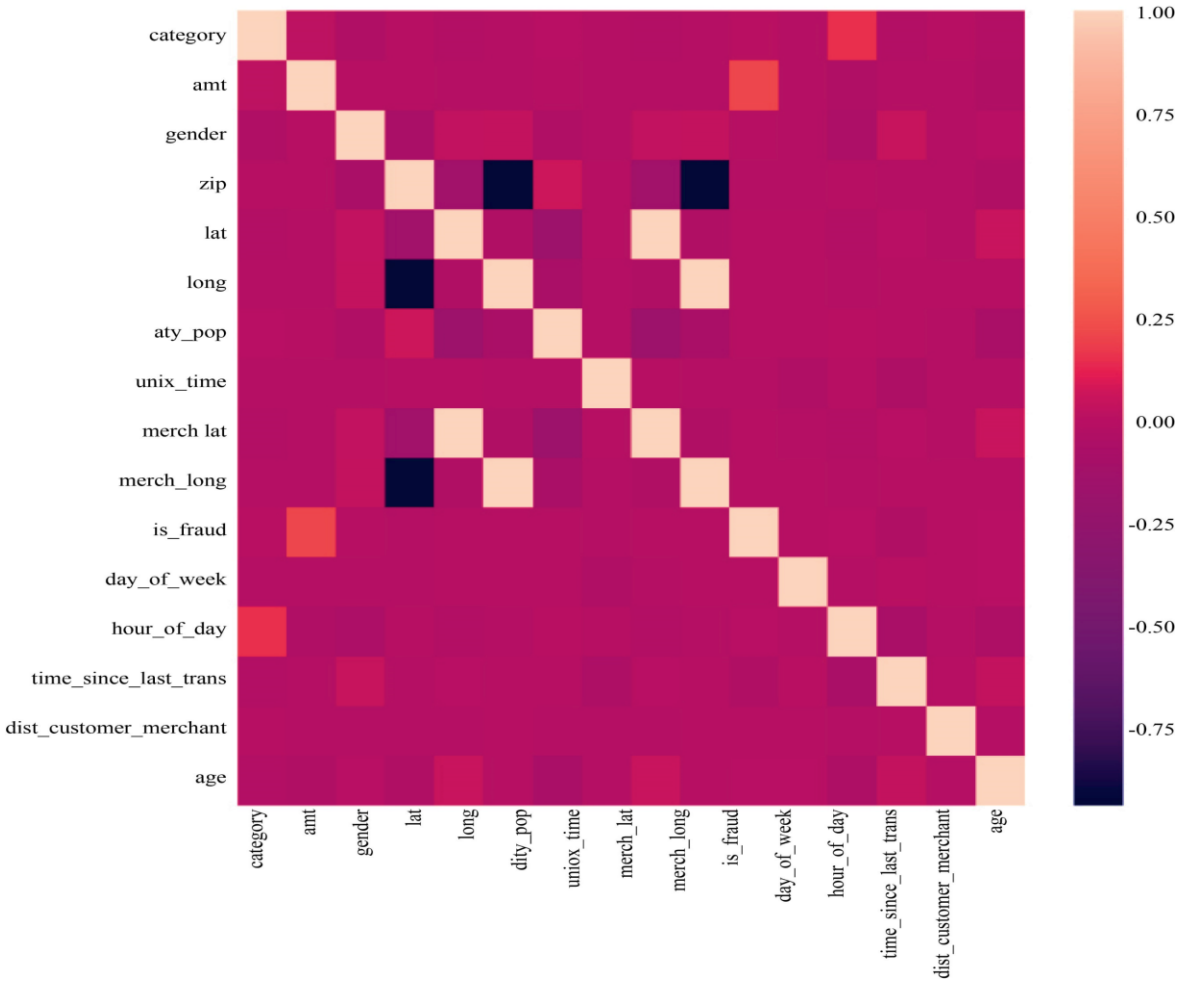

Fig. 9 Correlation heatplot

4. Results and Discussion

Various features of the data set have been analyzed, and several insights have been obtained. The 'trans_date_trans_time' feature has been broken down into several components like 'Age', 'day of the week', and 'month' in order to facilitate our analysis. These features have been thoroughly analyzed. It has been found that most transactions are being done after 12 noon and that during holiday seasons, as shown in Figure 4, the number of transactions along with the number of fraudulent transactions will increase. Also, old age people above 75 years are more susceptible to fraud, as shown in Figure 8. This is because fraudsters might try to take advantage of their lack of knowledge about the constantly changing ways how transactions are made.

The 'Female' gender people have been observed to do much of the transaction according to the dataset shown in Figure 7. Hence, transactions involved might be much more prone to fraud.

Similarly, several job profiles wherein a 100% fraud rate is seen can also be checked for discrepancies. In the categories feature, gas_transport, grocery_pos, home, shopping_pos, and kids_pets are the top 5 most categories with the most fraud rates. Most of these categories seem to involve either an online transaction sale or a POS sale. There might be some issues on those fronts, like tampered POS machines or hacked transaction gateway, which the company can check.

The processed dataset has been subjected to the implementation of three distinct algorithms. Three distinct sampling techniques have been employed to achieve dataset balance. The algorithms were applied to the dataset prior to its balancing for the purpose of demonstration.

Therefore, a total of twelve distinct models have been generated. Among the twelve (12) models constructed, the Random Forest Classifier, developed through the utilisation of the SMOTE sampling technique and subsequent hyperparameter tuning, has yielded the most desirable outcome, exhibiting an accuracy rate of 97.4%. Therefore, Random Forest Classifier - SMOTE sampling with hyperparameter tuning is the optimal model for predicting credit card fraud in work.

Table. 1 Findings of constructed twelve models

	Model Name	Training Score	Testing Score	Accuracy	F1 Score	Precision	Recall
0	Logistic Regression-Imbalance Class	0.993701	0.993747	0.993747	0.991093	0.000000	0.000000
1	Logistic Regression-with Random Undersampling	0.836315	0.835258	0.835258	0.833685	0.916207	0.738011
2	Logistic Regression - Random Over Sampling	0.837639	0.832594	0.832594	0.830971	0.913249	0.734772
3	Logistic Regression - SMOTE	0.835276	0.841486	0.841486	0.839747	0.890376	0.737319
4	Decision Tree - imbalance class	0.998622	0.954263	0.998026	0.997946	0.890376	0.757112
5	Decision Tree - Random Under Sampling	0.987438	0.954263	0.954263	0.954263	0.951370	0.957295
6	Decision Tree - Random Over Sampling	0.991297	0.946100	0.946100	0.946100	0.937265	0.955243
7	Decision Tree - SMOTE	0.994949	0.940217	0.940217	0.940217	0.950549	0.930108
8	Random Forest - imbalance class	0.999997	0.998257	0.998257	0.998147	0.954211	0.738731
9	Random Forest - Random Under-Sampling	1.000000	0.965808	0.965808	0.965803	0.977211	0.953737
10	Random Forest - Random Over Sampling	1.000000	0.954978	0.954978	0.954965	0.967148	0.941176
11	Random Forest - SMOTE	1.000000	0.948370	0.948370	0.948369	0.948122	0.949821
12	Random Forest - SMOTE [Hyperparameter Tuned]	1.000000	0.953804	0.973804	0.953805	0.955117	0.953405

5. Conclusion

This study aims to enhance comprehension of credit card fraud detection and its integration into the classification task to improve fraud detection rates. This study successfully utilized three classification algorithms, namely logic Regression, random forest algorithm and the Decision Tree classifier algorithm; three sampling techniques were used to balance the dataset. Twelve (12) different models were used for the prediction of credit card fraud. This present research employed a data-level methodology, incorporating various resampling techniques like undersampling, oversampling, and hybrid strategies. Accuracy, recall, precision, and f1-score were among the metrics calculated. The optimal model had a recall of 0.99, a f1 score of 0.99, and a precision of 0.99, and it was a Random Forest Classifier constructed using the SMOTE sampling technique after hyperparameter tuning. The outcome of this research will enhance the existing financial and card security technology, ultimately reducing losses. The implementation of this technology will enhance the ability of banks, card owners, and security agencies to promptly and accurately respond to incidents of theft or fraud.

However, the imbalance class may exhibit a tendency to display a bias towards the genuine ones. In the future, we may wish to investigate instances of credit card fraud based on location, transaction time and amount.

Acknowledgments

We want to acknowledge kaggle.com for the dataset used for this paper.

References

[1] Total Number of Cases of Online Banking Fraud in Belgium from 2006 to 2019, Statista, 2020. [Online]. Available: https://www.statista.com/statistics/614358/cases-of-internet-banking-fraud-in-belgium/

[2] Şerafettin Şentürk, Elif Yerli, and İbrahim Soğukpınar, "Email Phishing Detection and Prevention by Using Data Mining Techniques," *International Conference on Computer Science and Engineering*, pp. 707-712, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[3] Pooja Tiwari et al., "Credit Card Fraud Detection Using Machine Learning: A Study," *Artificial Intelligence, arXiv Preprint*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Market Share of Payment Methods in Total E-Commerce Transaction Value Worldwide in 2022, by Region, statista, 2022. [Online]. Available: https://www.statista.com/statistics/348004/payment-method-usage-worldwide/

[5] Caitlin Mullen, Card Industry's Fraud-Fighting Efforts Pay Off: Nilson Report, Payments Dive, 2023. [Online]. Available: https://www.paymentsdive.com/news/card-industry-fraud-fighting-efforts-pay-off-nilson-report-credit-debit/639675/

[6] Mohamed Ashraf, Mohamed A. Abourezka, and Fahima A. Maghraby, "A Comparative Analysis of Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques," *Digital Transformation Technology, Lecture Notes in Networks and Systems*, vol. 224, pp. 267-282, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Anuruddha Thennakoon et al., "Real-time Credit Card Fraud Detection Using Machine Learning," *9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 488-493, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Gaurav Kumar Singh et al., "Credit Card Fraud Detection Using Isolation Forest," *International Journal of Recent Advances in Multidisciplinary Topics*, vol. 2, no. 6, pp. 118-119, 2021. [Google Scholar] [Publisher Link]

[9] Kavya Gupta et al., "Learning Based Credit Card Fraud Detection-A Review," *International Conference on Applied Artificial Intelligence and Computing*, pp. 362-368, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Poonam M. Deshpande et al., "Applications of Data Mining Techniques for Fraud Detection in Credit-Debit Card Transactions," *National Conference on Technological Advancement and Automatization in Engineering*, 2016. [Google Scholar] [Publisher Link]

[11] Daniel Nduka Anowu et al., "Financial Forensic Analysis and Fraud Deterrence in Listed Deposit Money Banks in Nigeria," *Gusau Journal of Accounting and Finance*, vol. 2, no. 4, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Emmanuel Paul, In 2020, Nigeria Lost ₦5b to Fraud in 9 Months: What you Need to Watch Out for, Techpointafrica, 2021. [Online]. Available: https://techpoint.africa/2021/02/22/nigeria-lost-5b-fraud-2020/

[13] NIBSS Insight: Fraud in Nigerian Financial Services, NIBSS, 2021. [Online]. Available: https://nibss-plc.com.ng/nibss-insight-fraud-in-the-nigeria-financial-services/

[14] R. Saravanan, and P. Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification," *Second International Conference on Intelligent Computing and Control Systems*, pp. 945-949, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Vladimir Nasteski, "An Overview of the Supervised Machine Learning Methods," *Horizons*, vol. 4, pp. 51-62, 2017. [Google Scholar] [Publisher Link]

[16] Jason Bell, "What is Machine Learning?," *Machine Learning and the City: Applications in Architecture and Urban Design*, pp. 207-216, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yogesh Kumar, Sameeka Saini, and Ritu Payal, "Comparative Analysis for Fraud Detection Using Logistic Regression, Random Forest and Support Vector Machine, *SSRN*, pp. 1-6, 2020. [Google Scholar] [Publisher Link]

[18] Mahmudul Hasan et al., "Attack and Anomaly Detection in IoT Sensors in IoT Sites Using Machine Learning Approaches," *Internet of Things*, vol. 7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Pooja Bhati, and Manoj Sharma, "Credit Card Number Fraud Detection Using K-Means with Hidden Markov Method," *SSRG International Journal of Mobile Computing and Application*, vol. 2, no. 2, pp. 15-18, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[20] Rimpal R. Popat, and Jayesh Chaudhary, "A Survey on Credit Card Fraud Detection Using Machine Learning," *2nd International Conference on Trends in Electronics and Informatics*, pp. 1120-1125, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[21] Praveen Kumar Sadineni, "Detection of Fraudulent Transactions in Credit Card Using Machine Learning Algorithms," *Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, pp. 659-660, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Emmanuel Ileberi, Yanxia Sun, and Zenghui Wang, "A Machine Learning Based Credit Card Fraud Detection Using the GA Algorithm for Feature Selection," *Journal of Big Data*, vol. 9, no. 1, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Bandr Fakiha, "Forensic Credit Card Fraud Detection Using Deep Neural Network," *Journal of Southwest Jiaotong University*, vol. 58, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] John O. Awoyemi, Adebayo O. Adetunmbi, and Samuel A. Oluwadare, "Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis," *International Conference on Computing Networking and Informatics*, pp. 1-9, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[25] Vaishnavi Nath Dornadula, and S. Geetha, "Credit Card Fraud Detection Using Machine Learning Algorithms," *Procedia Computer Science*, vol. 165, pp. 631-641, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[26] Esraa Faisal Malik et al., "Credit Card Fraud Detection using a New Hybrid Machine Learning Architecture," *Mathematics*, vol. 10, no. 9, pp. 1480, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] J. Femila Roseline et al., "Autonomous Credit Card Fraud Detection Using Machine Learning Approach," *Computers and Electrical Engineering*, vol. 102, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Fawaz Khaled Alarfaj et al., "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," *IEEE Access*, vol. 10, pp. 39700-39715, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[29] Joy Iong-Zong Chen, and Kong-Long Lai, "Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert," *Journal of Artificial Intelligence*, vol. 3, no. 2, pp. 101-112, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Oded Maimon, and Lior Rokach, *Data Mining With Decision Trees: Theory and Applications*, World Scientific Publishing, 2014. [Google Scholar] [Publisher Link]

[31] Kartik Shenoy, Credit Card Transactions Fraud Detection Dataset, Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTrain.csv

[32] V. Rodriguez-Galiano et al., "Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines," *Ore Geology Reviews*, vol. 71, pp. 804-818, 2015. [CrossRef] [Google Scholar] [Publisher Link]